

Sobre errores y medición en las encuestas por muestreo

En estadística hablamos de robustez de un procedimiento cuando los supuestos que se tomaron en cuenta para desarrollar el mismo se pueden alterar sin que necesariamente sufran los resultados de la aplicación del mismo. Un ejemplo clásico es el intervalo confidencial para la estimación de la media poblacional de una población normal. Ese estimador es robusto, es decir, el procedimiento proporciona la cobertura esperada de la media poblacional aún cuando la distribución parental no sea gausiana.

No es descabellado preguntarnos si las encuestas por muestreo son instrumentos robustos en el sentido que le atribuimos al término en estadística. La justificación de tal pregunta es, creo, evidente. Mucho se discute en el momento presente en relación a cómo es posible que encuestas por muestreo que investigan el mismo fenómeno, por ejemplo, la preferencia de los votantes por un determinado candidato que aspira a un cargo de elección popular, produzcan resultados tan distintos. Si asumimos que los encuestadores responsables de estos estudios conocen su oficio y que obran éticamente, entonces habría que concluir que las encuestas no son procedimientos robustos, sino en el sentido estadístico estricto, al menos en un sentido cercano.

La Metodología de Encuestas, que ha resultado de una conjunción relativamente reciente de la Probabilidad, la Estadística, la Psicología Social y la Cognitiva, la Sociología, la Antropología y la Computación (Groves et al, 2004), proporciona una aproximación al tema de las encuestas por muestreo que ayuda a reflexionar en relación a la afirmación anterior. Esta aproximación se conoce como el Pa-

radigma del Error Total y básicamente se apropia del enfoque de la calidad total para estudiar los diferentes procesos involucrados en el diseño de una encuesta por muestreo.

El ciclo de vida de una encuesta por muestreo que se presenta en el gráfico 1, distingue diez componentes presentes en todas ellas (encerrados en rectángulos) que están insertos en una sucesión de etapas concatenadas. El paso de una a la siguiente ofrece oportunidades para la equivocación y el error, considerado éste como la discrepancia entre lo planificado y lo logrado. La búsqueda de la calidad en la encuesta impone que el investigador debe trabajar sobre cada oportunidad de error para disminuir su magnitud, sino eliminar su efecto. Los óvalos describen entonces conceptos de calidad a los que prestar debida atención.

El objetivo del diseño de encuestas por muestreo¹, es la minimización del error de la encuesta mediante la planificación y la elección de métodos de estimación que aseguren la reducción de las divergencias entre etapas sucesivas. El insumo de una fase es el resultado de la anterior, por lo tanto para que el error total permanezca dentro de límites razonables hay que evitar la amplificación del mismo en el paso de una fase a la siguiente.

Los componentes de calidad son propiedades de los estadísticos individuales de la encuesta y no de la encuesta como un todo. En el esquema, se empleó el estadístico media muestral como estimador del promedio de un constructo determinado en la población objetivo, pero se hubiera podido considerar cualquier otro estadístico muestral.

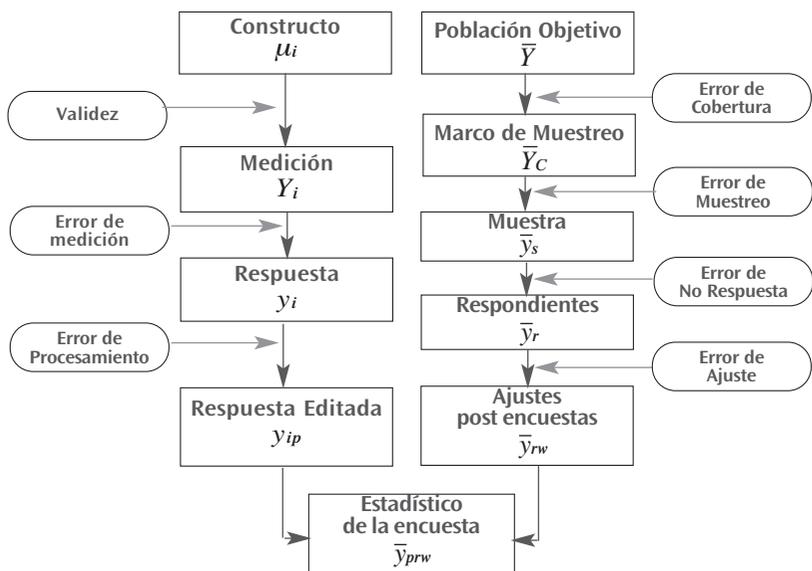
Este artículo contiene, palabras más, palabras menos, lo que debió ser mi intervención en el conversatorio sobre encuestas organizado por el Centro Gumilla y animado por los profesores Marcelino Bisbal y Pasquale Nicodemo.

En el texto se hacen reflexiones en torno al tema de los errores que se pueden cometer en el diseño y realización de las encuestas por muestreo y en los procesos de medición asociados.

■ ALBERTO CAMARDIEL

GRÁFICO 1:

Ciclo de Vida de una Encuesta por Muestreo desde LA PERSPECTIVA DE LA CALIDAD



Los símbolos empleados tienen el siguiente significado:

μ_i : Valor del constructo para el i -ésimo individuo ($i = 1 \dots N$) de la población

Y_i : Valor de la medición para el i -ésimo individuo ($i = 1 \dots \pi$) de la muestra

y_i : Valor de la respuesta de la medición para el i -ésimo individuo de la muestra

\bar{Y} : Media de la población objetivo

\bar{Y}_c : Media de la población marco

\bar{y}_s : Media de la muestra s -ésima

\bar{y}_r : Media de los respondientes en la muestra s -ésima

\bar{y}_{rw} : Media ajustada de los respondientes en la muestra s -ésima

El tránsito desde μ_i hasta y_{ip} se explica en términos del proceso de medición. El indicador Y_i es imperfecto, problemas en la administración de la medición hacen que obtengamos y_i en lugar de Y_i y la “corrección” de las debilidades de y_i nos lleva a y_{ip} .

En la columna de la izquierda vamos desde la media de la población objetivo \bar{Y} a la media de aquella parte de la población cubierta por el marco de muestreo \bar{Y}_c . Como únicamente se investiga a una parte de la población marco (muestra), sólo estamos en condiciones de calcular la media de la muestra seleccionada \bar{y}_s . Si además, no todos los individuos seleccionados responden, lo que puede ocurrir de forma total o parcial (responden algunas, pero no todas las preguntas), habrá que considerar la media muestral de los respon-

dientes \bar{y}_r , que a su vez requerirá de ajustes para tomar en cuenta los efectos de la no respuesta y de otras contingencias acaecidas y no previstas en el diseño y realización de la encuesta. El producto final de la encuesta, la media muestral \bar{y}_{prw} calculada sobre las respuestas y_{ip} incorporando todos los ajustes considerados, se emplea como un estimador de la media poblacional del marco. Los muestristas favorecen el uso de estimadores que tengan propiedades deseables como el insesgamiento y la consistencia².

La estimación válida y confiable de un parámetro poblacional, como por ejemplo, la proporción de personas que afirman que votarían por el candidato A de realizarse unas hipotéticas elecciones el próximo domingo, requiere que el instru-

mento de medición del constructo de interés sea válido, que la aplicación del mismo sobre los respondientes se realice sin error de medición, que las respuestas obtenidas eviten errores de procesamiento, que la población objetivo se represente adecuadamente mediante un marco de muestreo completo y actualizado, que el error de muestreo se pueda calcular con los datos de la muestra, es decir que la muestra sea medible en el sentido de Kish³ y se mantenga dentro de límites razonables, que todos los respondientes hayan respondido y que no se hayan presentado situaciones que requieran de ajustes posteriores en la encuesta.

Pero todos los que hemos trabajado alguna vez con encuestas en la investigación social, sabemos que estos buenos deseos que el investigador incorpora en su diseño, rara vez, por no decir nunca, se cumplen en la práctica.

En su lugar, tenemos que manejarnos con instrumentos de medición de constructos muy abstractos con validez menor que la deseada, instrumentos que producen un error de medición que no se puede ignorar y por lo tanto de fiabilidad limitada, errores de procesamiento muy humanos y por eso difíciles de eliminar totalmente, marcos de muestreo imperfectos y posiblemente no muy actualizados que determinan una cobertura parcial de la población objetivo. La propia naturaleza de las encuestas por muestreo supone un error ineludible por el hecho de sólo incluir a una parte de la población (aunque quizás éste sea el que menos preocupe a los estadísticos porque es el que mejor se conoce y se puede medir y manejar eficientemente). La no respuesta total o parcial es también imposible de evitar y sabemos que su presencia puede atentar de forma drástica en contra de la capacidad de representación que tiene una muestra probabilística y como consecuencia de todo lo descrito se requiere de operaciones de ajuste post-encuesta que también traen consigo la posibilidad de cometer errores asociados con estos procedimientos.

Repetimos que sólo podemos cuantificar el error de muestreo en muestras probabilísticas. Muchos de los procedimientos empleados en la práctica cotidiana de las encuestas, no cumplen con tal condición. Algunos de ellos emplean múltiples etapas de selección, en las que una o más pueden ser probabilísticas y una o más de las restantes son intencionales, pero el resultado final no es una muestra probabilística⁴.



El esquema considerado permite también destacar un aspecto que no siempre es explícitamente reconocido cuando hablamos y pensamos en relación a las encuestas por muestreo, a saber, que en este tipo de investigación están presentes dos modos de inferencia⁵. La columna izquierda del gráfico 1 corresponde a lo que se denomina la inferencia de la medición y la derecha a la inferencia de la representación. El primer tipo de inferencia nos lleva de las respuestas proporcionadas por los entrevistados hasta las características de los mismos y el segundo nos permite pasar desde las características de la muestra a las características de la población. Ambas se encuentran en el estadístico de la encuesta que cuando consideramos la estimación de la media poblacional, denotamos por \bar{y}_{prv} .

En nuestro ejemplo de votación, el estadístico de la encuesta es el porcentaje de respondientes que expresaron que sufragarían a favor del candidato A de llevarse a cabo las elecciones el próximo domingo⁶. Este estadístico que podemos denotar de forma simplificada por p es un estimador del parámetro poblacional correspondiente P . El estadístico p estará afectado en mayor o menor grado por las fuentes de error antes identificadas. Por ejemplo, si la muestra se toma en ciudades de 20.000 habitantes o más, el error de cobertura producirá un sesgo considerable de aproximadamente un 30% si nos atenemos a las cifras proyectadas con los datos del censo de población y vivienda del año 2001. Pero en el reporte de los resultados de las empresas encuestadoras locales que conocemos (o en sus fichas técnicas) se presenta la magnitud del error de muestreo que sería aplicable, de tomar en cuenta la incertidumbre encontrada en la investigación y que eventualmente se pudiera emplear, por ejemplo para dar estimaciones por intervalo confidencial⁷ del parámetro poblacional P . Este énfasis en el error de muestreo, ni es correcto en muchos casos porque la muestra no es probabilística, ni es la única fuente de error a considerar, ni tampoco es, posiblemente, el que aporta mayor cuantía al error total.

Estas dos formas de inferencia son esenciales para que las encuestas por muestreo funcionen adecuadamente, porque proporcionan las bases sobre las que se construye la estructura, a saber, que:

1. Las respuestas dadas por los encuestados describan adecuadamente las características de los mismos.
2. El subconjunto de las personas que participan en la encuesta deben tener

(...) cuando tratamos con encuestas por muestreo de poblaciones humanas con el fin de estudiar la realidad social o política, la mirada debe dirigirse hacia la psicología. Los estudios de mayor relevancia para la medición de lo social provienen de los avances logrados por la psicometría a lo largo del siglo pasado

características semejantes al conjunto de los individuos de la población objetivo que se pretende estudiar.

Aunque las razones por las que los resultados de las encuestas que estudian un mismo fenómeno, por ejemplo el electoral, terminan con cifras muy discrepantes, que se originan en la falta de control de uno o más de los componentes del ciclo de vida de la encuesta descrito en el gráfico 1, nuestra sospecha es que mayormente están asociadas con los procesos pertenecientes a la inferencia de la medición. Y esto nos lleva a considerar el segundo aspecto importante sobre el que trataremos en este trabajo, la medición en las encuestas por muestreo.

Los estadísticos hemos considerado tradicionalmente que el asunto de la medición compete a otros o cuando mucho que lo único que deberíamos saber, es si los datos que nos han proporcionado terceras personas para su análisis, vienen medidos en escala nominal, ordinal, de intervalo o de razón. Esta conducta deviene del trabajo de un notable psicólogo norteamericano del siglo pasado, Stanley Smith Stevens, que quizás, como derivación de sus trabajos en psicofísica defendió que era posible medir en las ciencias sociales y humanas de la misma forma como se mide en las ciencias físicas. Los trabajos de Stevens fueron muy influyentes en la comunidad estadística y dieron origen a una elaboración muy trabajada por parte de la filosofía de la ciencia, que se ha compendiado en lo que se conoce como la escuela representacional de la medición. Pero esto es filosofía y de poca ayuda en la solución

de los problemas prácticos de la medición en las ciencias sociales.

La relación de los estadísticos con la medición ha venido cambiando con el tiempo. Por ejemplo, David Hand defiende que la medición debería ocupar en la estadística un lugar tan destacado como el que tiene la probabilidad (Hand, 1996; Pág. 445). Otro estadístico contemporáneo, David Bartholomew considera que la medición, bien sea de características poblacionales (por ejemplo la tasa de desempleo en el país) o individuales (por ejemplo la actitud de las personas ante una medida como la anunciada recientemente de eliminar los antecedentes policiales de personas que hayan cumplido sentencias por delitos cometidos) es un problema de inferencia estadística, en el primer caso de estimación de parámetros y en el segundo, de predicción de valores de una variable aleatoria. En ambas situaciones es necesario recurrir al uso explícito de modelos estadísticos que incorporen componentes sistemáticos y aleatorios (Bartholomew, 1996).

Pero cuando tratamos con encuestas por muestreo de poblaciones humanas con el fin de estudiar la realidad social o política, la mirada debe dirigirse hacia la psicología. Los estudios de mayor relevancia para la medición de lo social provienen de los avances logrados por la psicometría⁸ a lo largo del siglo pasado, por ejemplo los conceptos de validez y de fiabilidad que describimos a continuación.

La divergencia entre el constructo y su medida asociada corresponde al campo de la validez. En Psicometría, se define la validez de constructo como la cuantía en la que la medida está relacionada con el constructo subyacente. Para la obtención de una expresión matemática de la validez, consideremos que se puede repetir hipotéticamente el proceso de medición del constructo μ_i un número infinito de veces. Entonces, el resultado de la medición t -ésima se puede representar mediante la siguiente expresión:

$$Y_{it} = \mu_i + \varepsilon_{it}$$

Esta ecuación afirma que las desviaciones de la medición de la respuesta obtenida para la encuesta realizada (Y_{it}) con respecto al valor verdadero (μ_i) varían de encuesta a encuesta. Ciertamente, realizamos la encuesta una sola vez, no muchas veces, pero consideramos que, conceptualmente, de haberse realizado, el respondiente se expresaría posiblemente de forma distinta en cada aplicación independiente de la encuesta.

La validez se define entonces como la correlación entre la medición Y_i y el constructo μ_i considerada sobre todas las posibles encuestas (pruebas) y personas:

$$E_{it} [(Y_{it} - \bar{Y}) (\mu_i - \bar{\mu})] / \sqrt{E_{it} (Y_{it} - \bar{Y})^2} \sqrt{E_i (\mu_i - \bar{\mu})^2}$$

en donde μ e \bar{Y} representan la media de los μ_i y de los Y_i respectivamente y el símbolo E_x denota el valor esperado o esperanza matemática sobre todos los x posibles.

Si la medición y el constructo covarían, entonces la medición tiene una elevada validez de constructo. Una medida válida de un constructo es aquella que está perfectamente correlacionada con el mismo.

El error de medida en una encuesta por muestreo se manifiesta por discrepancias entre el valor verdadero de la medición cuando se aplica a un elemento de la muestra y el respectivo valor obtenido en la muestra. Es usual distinguir entre el sesgo de respuesta y la varianza de respuesta.

Si para el individuo i -ésimo denotamos la primera cantidad por Y_i y la segunda cantidad por y_i , entonces la diferencia:

$$y_i - Y_i$$

representa la desviación de respuesta para ese individuo.

En el caso que todas las desviaciones de respuesta se presenten en una misma dirección, medimos el sesgo de respuesta cuando se consideran las infinitas “pruebas” hipotéticas (aplicaciones de la medida) de la forma siguiente:

$$E_t(y_i) - Y_i$$

Otro tipo de error de respuesta que también está presente en el proceso de medición se expresa por la inestabilidad en las respuestas cuando se consideran las infinitas e hipotéticas aplicaciones de la medida. Si esta variabilidad es alta se dice que la respuesta tiene baja confiabilidad o fiabilidad. A esta variabilidad se le denomina varianza de respuesta.

A partir de los años ochenta del siglo pasado, otra fuente de importantes avances y aportes a la medición la proporcionó la psicología cognitiva. El proceso de formular preguntas y responderlas se ha estudiado desde esta disciplina mediante el uso de modelos, que si bien es cierto no han generado un conjunto de pautas universales para la elaboración del cuestionario ideal, si que han permitido organizar el conjunto de hallazgos desconectados que se había venido acumulando con anterioridad en relación, por ejemplo, a

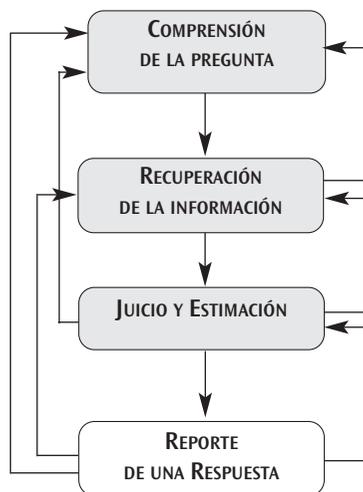
las respuestas de preguntas factuales versus preguntas de opinión, a las respuestas de preguntas abiertas versus preguntas cerradas, a la influencia del formato de las opciones en las preguntas cerradas y al efecto del orden de las preguntas en el cuestionario.

La mayoría de los modelos que tratan de explicar el proceso de respuesta en las encuestas incluyen cuatro grupos de procesos: (1) comprensión, (2) recuperación o evocación, (3) juicio y (4) reporte. Nos referimos a la comprensión de la respuesta formulada, la recuperación de la información necesaria para responder la pregunta, al juicio y la estimación en la que se combina o resume la información recuperada y al reporte en el que verbaliza la respuesta en el formato requerido. Los procesos y sus posibles interacciones se representan en el gráfico 2 (Groves et al, 2004; Pág. 202).

En ocasiones puede ser importante tomar en cuenta el proceso cognitivo que tuvo lugar antes de la entrevista y que se refiere a los eventos experimentados por los que se pregunta. Este proceso se denomina codificación y se refiere a la formación de la memoria a partir de la experiencia. Existe evidencia que se puede mejorar las preguntas en las encuestas si se toma en cuenta como codificó el respondiente la información que pretende evocar.

GRÁFICO 2:

Modelo del proceso de respuestas en encuestas por muestreo



Con frecuencia los respondientes saltan alguno de los subprocesos y realizan otros de forma desatenta o despreocupada. En general, las preguntas que se formulan en las encuestas por muestreo son difíciles y exigen mucho de la memoria o requieren la elaboración de juicios complicados. La mayoría de los respondientes proporcionan probablemente sólo estimaciones gruesas como respuestas. No existe razón para suponer que los respondientes típicos tengan el tiempo o la disposición o la dedicación para trabajar tenazmente en la construcción de la respuesta, además existe amplia evidencia de que los respondientes emplean atajos para simplificar su tarea (Ibíd. Pág. 203).

Los respondientes no llevan a cabo necesariamente los procesos cognitivos relevantes en una sucesión fija comenzando con la comprensión y terminando con el reporte. Con seguridad hay un ir y venir y un cierto solapamiento entre subprocesos.

El supuesto básico del análisis cognitivo del proceso de responder en las encuestas por muestreo, es que los fallos en las operaciones cognitivas involucradas en la producción de respuestas son los principales responsables de los errores en las respuestas. Se distinguen siete problemas:

1. Fallos en la codificación de la información buscada.
2. Interpretación errónea de las instrucciones.
3. Olvido y otros problemas de la memoria.
4. Juicios o estrategias de estimación defectuosos.
5. Problemas en el formateo de las respuestas.
6. Reporte más o menos falso de forma deliberada
7. Fallos en el seguimiento de las instrucciones

Michael Ornstein culmina un escrito sobre diseño de cuestionarios, afirmando que la mejor investigación que emplea encuestas por muestreo, es la realizada por personas que saben mucho más que sus antecesores acerca del diseño de encuestas, pero que combinan ese conocimiento con habilidades artesanales para generar preguntas que la gente entienda y que sirvan para describirlos (Ornstein, 1998; Pág. 44). Afirma además que cuando las

preguntas están cuidadosamente concebidas, se logra dar cuenta entre la mitad y dos tercios de la varianza del error de medición. El resto depende de otras fuentes de errores muy variables como por ejemplo:

1. La interpretación adecuada de las preguntas.
2. El estado de ánimo de los respondientes.
3. Los eventos acontecidos en el último momento.
4. La percepción que tiene el respondiente de la entrevista, si ese fuera el caso.
5. El impacto que han podido ejercer las preguntas anteriores.

Después de este largo periplo por el territorio del error en las encuestas por muestreo, cabe preguntarse si a pesar de todo, vale la pena emplear encuestas por muestreo en la investigación social empírica. Los indudables aciertos acumulados por más de 70 años de práctica continuada de esta forma de investigación social hablan por sí solos, si vale la pena cuando se trabaja con honestidad (ética), conocimiento (teoría) y ajustados a la mejor experiencia verificada (práctica validada). Una explicación posible de por qué funcionan se encuentre en la especulación final de Ornstein: quizás opere en las encuestas por muestreo, un teorema análogo al teorema central del límite, en el que las fuentes de error se cancelan entre sí, contribuyendo al error de la medición pero no al sesgo.

ALBERTO CAMARDIEL

Profesor Titular de la UCV. Estadístico y Profesor del Postgrado en Estadística y Actuariado, UCV, FACES.

Referencias

- AZORIN, F. (1962): *Curso de muestreo y aplicaciones*. Segunda edición. Madrid: Presidencia del Gobierno, Instituto Nacional de Estadística.
- BARTHOLOMEW, D.J. (1996): *The statistical approach to social measurement*. San Diego: Academic Press.
- GROVES, R.M., F. Fowler Jr., M. Couper, J. Lepkowski, E. Singer y R. Tourangeau. (2004): *Survey methodology*. New Jersey: Wiley-Interscience.
- HAND, D.J. (1996): "Statistics and the Theory of Measurement". En: *Journal of the Royal Statistical Society, Series A*, Vol. 159, Part 3, pp. 445-492.

Notas

- 1 Para Leslie Kish, el diseño de la muestra incluye dos asuntos cruciales: el proceso de selección que son las reglas y operaciones mediante las cuales los elementos de la población se incluyen en la muestra (por ejemplo métodos EPSEM o de igual probabilidad de selección y No-EPSEM, muestreo de elementos y muestreo de conglomerados, selección estratificada, selección aleatoria y selección sistemática, selección en una fase y en dos o más fases) y el proceso de estimación mediante el cual se calculan los estadísticos muestrales que son las estimaciones muestrales de los valores poblacionales (Kish, 1965; Pág. 4).
- 2 Se dice que un estimador y_n es insesgado cuando su valor esperado coincide con el parámetro que se desea estimar, en símbolos, $E(y_n) = Y$. Decimos que y_n es consistente cuando el estimador se acerca al valor poblacional Y a medida que aumenta el tamaño de la muestra, esto es: $\lim_{n \rightarrow \infty} Pr\{|y_n - Y| \geq \epsilon\} = 0$ para todo $\epsilon > 0$.
- 3 Para que una muestra sea medible debe ser probabilística y tal que su diseño permita que la inferencia estadística de la muestra a la población muestreada se base en medidas de variabilidad, en general errores estándar, calculados a partir de los datos muestrales (Kish, 1965; Pág. 20). Una muestra es probabilística si la probabilidad de selección de cualquier elemento de la población en la muestra es conocida y positiva.
- 4 Azorín califica de muestras mixtas aquellas muestras que pretenden ser probabilísticas pero que tienen ciertos componentes de selección intencional. Menciona este autor que Cochran, Mosteller y Tukey en su crítica al Reporte Kinsey, hablan de muestreo semiprobabilístico superior cuando se conoce la probabilidad de extracción de un segmento de la población, pero no se conoce la probabilidad de selección de un elemento dentro de él. Y hablan de muestreo probabilístico inferior cuando se conoce la probabilidad de selección de elementos dentro del segmento, pero no se conoce la probabilidad de selección del segmento. El muestreo es probabilístico cuando es semiprobabilístico superior y semiprobabilístico inferior (Azorin, 1962; Pág. 6).
- 5 Empleamos el término inferencia en el sentido de describir fenómenos no observados a partir de fenómenos observados.
- 6 Cuando codificamos la no posesión de un atributo dicotómico por 0 y su posesión por 1, la media de ceros y unos para un determinado conjunto de observaciones es la proporción de casos que poseen el atributo.
- 7 Una estimación por intervalo confidencial de un parámetro poblacional P , es una declaración probabilística de la forma $Pr\{l_i \leq P \leq l_s\} = 1 - \alpha$, en donde los límites inferior y superior se definen como $l_i = p - t_{\alpha} * (Error\ estándar\ de\ p)$ y $l_s = p + t_{\alpha} * (Error\ estándar\ de\ p)$. Esta afirmación se verbaliza corrientemente, declarando que el parámetro poblacional P está comprendido entre l_i y l_s con una confianza del $(1 - \alpha) 100\%$.
- 8 El sociólogo Otis Duncan afirma, no sin críticas, que la psicometría ha resultado para las ciencias sociales en "... un paradigma metodológico y una fuente de técnicas específicas de medición ..." (Duncan, 1984; Pág. 203).

KISH, L. (1965): *Survey sampling*. New York: John Wiley & Sons.

ORNSTEIN, M. (1998): "Questionnaire Design." En: *Current Sociology*, Vol. 46, #4, pp. 7-47.

SEIJAS, F.L. (1993): *Investigación por muestreo*. Segunda edición revisada y ampliada. Caracas: UCV, FACES, Unidad de Publicaciones y Reproducción.

DUNCAN, O.D. (1984): *Notes on social measurement: historical & critical*. New York: Russell Sage Foundation.